Recurrence plot analysis of irregularly sampled data

Ibrahim Ozken,[1] Deniz Eroglu,[2,3] Sebastian F. M. Breitenbach,[4] Norbert Marwan,[2] Liangcheng Tan,[5] Ugur Tirnakli,[1] and Jürgen Kurths[2,3,6]

[1]*Department of Physics, Faculty of Science, Ege University, 35100 Izmir, Turkey*

[2]*Potsdam Institute for Climate Impact Research, 14473 Potsdam, Germany*

[3]*Department of Physics, Humboldt University, 12489 Berlin, Germany*

[4]*Sediment and Isotope Geology, Institute for Geology, Mineralogy & Geophysics, Ruhr-Universität Bochum, Universitätsstraße 150, 44801 Bochum, Germany*

[5]*State Key Laboratory of Loess and Quaternary Geology, Institute of Earth Environment, Chinese Academy of Sciences, Xi'an 710061, China*

[6]*Department of Physics, Saratov State University, Astrakhanskaya Street 83, 410012 Saratov, Russia*

(Received 7 September 2018; published 15 November 2018)

Irregularly sampled time series usually require data preprocessing before a desired time-series analysis can be applied. We propose an approach for distance measuring of pairs of data points which is directly applicable to irregularly sampled time series. In order to apply recurrence plot analysis to irregularly sampled time series, we use this approach and detect regime transitions in prototypical models and for an application from palaeoclimatatology. This approach might be useful for any method that is based on distance measuring, e.g., correlation sum or Lyapunov exponent estimation.

DOI: 10.1103/PhysRevE.98.052215

I. INTRODUCTION

One of the main goals of data science is to understand dynamical processes and to predict their future in order to evaluate vulnerability and resilience, e.g., to prepare the society for recurring events such as natural catastrophes and economical or medical crises. For this purpose, it is indispensable to understand the past with a complete range of natural variability, which is in general obtained from time series. Time series represent the variation of dynamical systems and provide, e.g., information on long-term trends but also abrupt and critical regime transitions (such as crises and catastrophes). However, not all of this information is easily accessible.

Besides linear methods, several techniques such as the Lyapunov exponent, Poincaré recurrences [1], or correlation dimension [2] have been proposed for a better understanding of the dynamics as represented by time series. Most of these methods are based on investigating the evolution of trajectories in the phase space or, where required, in the reconstructed phase space [3]. The standard procedure of analyzing a dynamical system by its phase space is by measuring the distances between points on the phase-space trajectory (or its change) and quantifying them. While numerous ways of measuring these distances are at our disposal, most of them require regular temporal sampling (i.e., regularly sampled time series), where the temporal resolution $\Delta t = t_i - t_{(i-1)} = \text{const}\, \forall i \in [1, N]$ (N is the total number of points in the time series). Unfortunately, it is not always possible to record data regularly, such as in astrophysics or in earth sciences. In the latter, for example, regular spatial sampling of palaeoclimate archives (e.g., lacustrine sediments or cave carbonates) translates into irregularly spaced sampling intervals in the time dimension due to varying deposition rates. Standard methods for distance measuring, such as Euclidean norm and maximum norm, are not directly applicable and require data preprocessing, such as interpolation, to regularize the investigated time series. Data preprocessing usually introduces additional uncertainty or bias [4].

A recently introduced data preprocessing approach to regularize time series is the transformation cost time series (TACTS) approach. It transforms irregularly sampled time series into regular ones [5]. The base of the TACTS method is a metric distance for the transformation of a certain block of the time series to match a subsequent block. In other words, TACTS is a kind of high-pass filter for preprocessing data. Here we will show that this distance measure is not only useful as a preprocessing step for regularization of data sets but also suggestive for a general distance measure in the phase space. The distance can be used in any analysis or employed by any method such as the recurrence plot (RP) and recurrence quantification analysis (RQA) [6]. This metric distance was adopted from [7] and extended in our previous work [5]. Suzuki *et al.* also used a similar distance measure based on the point process to construct the RP [7]. In this study, we focus on RPs and RQA, but again, our approach is not limited to these methods; it can be applied in any technique where measuring the distance is required, especially in the case of irregular sampling. Combining this specific distance measure approach with RPs, we identify regime transitions in dynamics in irregularly sampled time series without the need for any preprocessing.

Our paper is organized as follows. In the following section we introduce our metric distance, the RP, and explain how to implement this measure into the RP. In Sec. III we apply this technique to paradigmatic model examples, the logistic map and the Rössler attractor, to evaluate the performance of our method on perturbed data by removing random points from the system and adding noise. Finally, as a real-world

application, we analyze a palaeoclimate record in Sec. IV. We chose the palaeoclimatic reconstruction ($\delta\,^{18}$O variations in a stalagmite) from Dayu Cave in China [8] as an illustrative test case for our method, since the interpretation of this geochemical proxy reconstruction has uniquely been linked to historical events (inscriptions on the cave's walls). The observed $\delta\,^{18}$O variation in the robustly radiometrically dated stalagmite tracks changes in summer monsoon rainfall and indicates the occurrence of massive droughts over the past 700 years. The inscriptions found in the same cave report severe water shortage in the region around the cave at exactly the same time as indicated by the speleothem proxy record. This corroboration of proxy data by historical information allows highly robust tests of the methodology proposed in this study.

II. TIME-SERIES ANALYSIS

A. Metric analyses

Measuring distance between state vectors in a phase space is often used to describe the behavior of dynamical systems. The evolution of the distance between trajectories of a dynamical system provides valuable information for the general behavior of systems such as their dynamics (determining the Lyapunov exponent or correlation dimension) [2,9] or their collective behavior [10,11]. Numerous methods are available to measure distances in phase space, such as Euclidean or maximum norm or, more specifically, order patterns [12]. Such distance measures require equidistant time points along the phase-space trajectory, a constant temporal sampling. However, observations in various fields are not always regularly sampled. The most common way to cope with these irregularities is by time-series interpolation. While interpolation anchors the sampling rate, it not only fills the gaps in a given time series, but also changes the position of all data points. Such an interpolation strategy can easily bias the results and cause more additional uncertainty in the interpolated data [13].

To address the inapplicability of interpolation approaches and the importance of finding a distance definition for irregularly sampled time series that were highlighted in a case that involved firing neurons, a metric distance was proposed by Victor and Purpura [14]. The main idea of this distance measure is a transformation of a data pattern to another one with two elementary operations: (i) adding or deleting a point and (ii) shifting a point in time. These operations have their defined costs and, crucially, the total cost of the transformation has to be minimal. Suzuki *et al.* adopted this measure and extended it for marked data [7] with arbitrary constants. Finally, Ozken *et al.* proposed the meaning of the arbitrary constants by using time series' properties [5]. The distance definition reached its final form to calculate the cost D between data segments S_a and S_b as

$$D(S_a, S_b) = \overbrace{\sum_{(\alpha,\beta)\in C} \left\{ \Lambda_0 \|t_a(\alpha) - t_b(\beta)\| + \frac{1}{m}\sum_{k=1}^{m} \Lambda_k \|L_a(\alpha) - L_b(\beta)\| \right\}}^{\text{shifting and amplitude change}} + \underbrace{\Lambda_S(|I| + |J| - 2|C|)}_{\text{adding and deleting}}, \tag{1}$$

where C is a set of pairs of events, time instances (α, β) corresponding to two different data segments that will be transformed from S_a to S_b.

The first term on the right-hand side of Eq. (1) sums the costs of shifting in time and changes in amplitude from the αth event in S_a to the βth event in S_b. The times of these points are denoted by $t_a(\alpha)$ and $t_b(\beta)$ and the amplitudes by $L_a(\alpha)$ and $L_b(\beta)$ for the αth and βth elements of S_a and S_b, respectively. In Eq. (1), m is the dimension of the phase space and $\|\cdot\|$ is a norm. The cost of shifting points in time is given by the factor Λ_0 and of changing amplitudes by Λ_k. These factors depend on the data and are given by

$$\Lambda_0 = \frac{M}{\text{total time}}, \tag{2a}$$

$$\Lambda_k = \frac{M-1}{\sum_i^{M-1} \|x_i - x_{i+1}\|}, \tag{2b}$$

where M is the total number of events and x_i is the amplitude of the ith point in the data.

In the second term on the right-hand side of Eq. (1), I and J are sets of indices of the events in S_a and S_b, respectively. In addition, $|\cdot|$ denotes the cardinality of sets. The cost of the adding or deleting a process is given by Λ_S and is used as an optimization parameter. The selection routine for Λ_S is the following: We calculate the distances between all data segments for the entire range of $\Lambda_S \in [0, 4]$ with the step size $\Delta\Lambda_S = 0.01$. Since each distance value is independent of the others, we expect to have normally distributed distances. Therefore, the optimal Λ_S is chosen by the best fit to a normal distribution.

It is worth noting that D is a metric distance. Intrinsically it satisfies the three essential conditions: $D(S_a, S_b) \geqslant 0$ (positive), $D(S_a, S_b) = D(S_b, S_a)$ (symmetric), and $D(S_a, S_c) \leqslant D(S_a, S_b) + D(S_b, S_c)$ (triangle inequality).

Now we illustrate how to find the metric distance D for an example (Fig. 1). An irregularly sampled time series $\{x_i\}_{i=1}^{M}$ is entirely divided by equally sized small sequences and we consider two of them which are given by the state a ($S_a = \{a_\alpha\}_{\alpha=1}^{4}$) and the state b ($S_b = \{b_\beta\}_{\beta=1}^{3}$). With the first and second steps, the events are shifted in time and naturally the amplitudes are changed. In the third and fourth steps points are deleted, while in the last a point is added. The cost of each step is shown in detail in Fig. 1 and the total cost of all steps is equal to the metric distance $D(S_a, S_b)$.

The distance is an essential measure in numerous mathematical techniques. Without harmful preprocessing techniques such as interpolation, the proposed metric distance approach can be directly implemented in such methods.

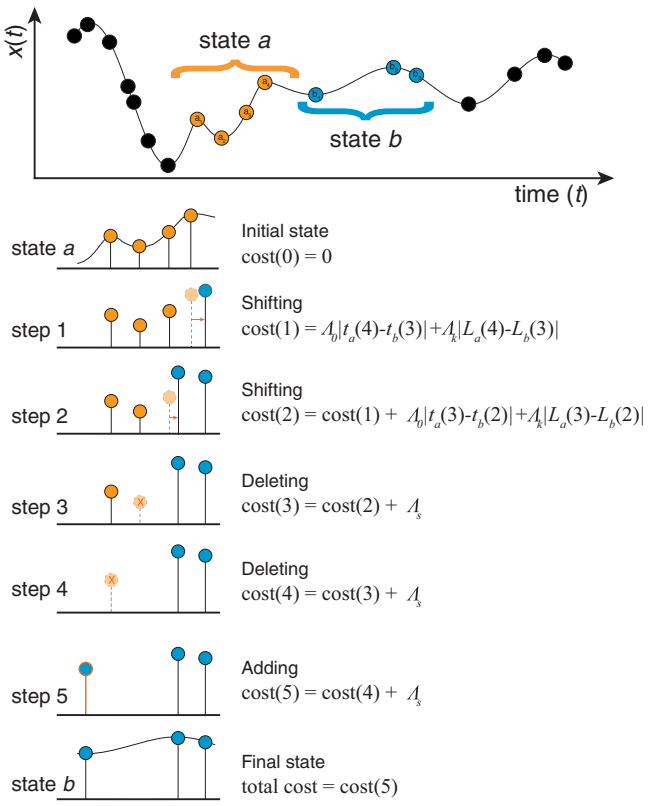

FIG. 1. Illustration of the transformation from state a to state b. In total state a undergoes five steps numbered as steps $1, 2, \ldots, 5$. Note that after the step 5 state a is transformed to state b. The path shown is a minimal-cost path and all the steps are elementary steps, like moving an event or deleting and creating. On the right-hand sides of the steps, the cumulative costs are given.

B. Recurrence plot

The RP approach is a versatile tool to study the behavior of dynamical systems [6] and was used to tackle many problems in several disciplines, e.g., electrochemistry [15], earth science [16], econophysics [7], and engineering [17]. Here we apply the RP for detecting regime transitions in prototypical dynamical systems and a real-world application on paleoclimate records.

The RP was introduced by Eckmann *et al.* as a matrix of pairwise Poincaré recurrences of phase-space states [1]. According to the Poincaré recurrence theorem, a trajectory $\vec{x}(t)$ of an m-dimensional system, in sufficiently long and finite time, will return to the ε neighborhood of a previous state [18]. The RP is a tool to visualize this recurrences with a time vs time matrix. For a given trajectory $\vec{x}_i$ ($i = 1, 2, \ldots, N$ and $\vec{x}_i \in \mathbb{R}^m$), the RP is defined as [6]

$$R_{i,j}(\varepsilon) = \Theta(\varepsilon - ||\vec{x}_i - \vec{x}_j||), \quad i, j = 1, \ldots, N, \qquad (3)$$

where $\Theta(\cdot)$ is the Heaviside function and $\|\cdot\|$ is a norm. Therefore, $R_{i,j} \equiv 1$ if the state at time i recurs to a former (or later) state at j, and $R_{i,j} \equiv 0$ otherwise. The selection of the recurrence threshold ε depends on a specific research question [19]: whether, in order to distinguish different dynamics, a threshold relative to the standard deviation of the time series is suitable [20].

The RP matrix is binary and symmetric and obviously the main diagonal of this matrix is always $R_{i,i} \equiv 1$. According to the behavior of the system, the recurrences create some patterns in the RP [6]. For instance, long, continuous diagonal lines appear in periodic systems, whereas RPs of chaotic systems have rather short diagonal lines and RPs of stochastic systems have no (or only very few and very short) diagonal lines. Such lines are used for subsequent quantification of the recurrence structure of the system (discussed below). However, defining and measuring such diagonals in irregularly sampled RPs is an unsolved problem. The standard approach is to interpolate the time series before the RP analysis. In contrast to this standard approach, we now implement our cost distance approach in the RP [Eq. (1)].

In order to apply the distance, we divide the time series into ω small sequences. From now on, the time indices will be the center points of these small sequences and the RP elements will represent the distance of the states. Therefore, the size of the RP matrix will be $\omega \times \omega$,

$$R_{i,j}(\varepsilon) = \Theta(\varepsilon - D(\vec{x}_i, \vec{x}_j)), \quad i, j = 1, \ldots, \omega, \qquad (4)$$

where i and j are the indices of the states as defined by the small sequences and D is the transformation distance between those states [Eq. (1)]. For this special variant of RP, it turns out that an alternative selection method for the recurrence threshold ε would be more appropriate than the standard ones. We compute the standard deviation σ_D of the distances $D(\vec{x}_i, \vec{x}_j)$ and define our threshold as $\varepsilon = 2\sigma_D$. In the following sections we always use this threshold selection method.

The quantification of the dynamics of dynamical systems from time series by RPs is well studied and several measures of complexity were introduced as recurrence quantification analysis [6,21]. For example, the frequency distribution of diagonal lines $P(\ell)$ is used for some of these measures since the length of the diagonal lines is related to the divergence behavior of the dynamics and thus indirectly related to the Lyapunov exponent of the system. One important measure of the RQA is determinism (DET), which quantifies the fraction of recurrence points ($R_{i,j} = 1$) which form diagonal lines,

$$\mathrm{DET} = \frac{\sum_{\ell=\ell_{\min}}^{N} P(\ell)}{\sum_{\ell=1}^{N} P(\ell)}, \qquad (5)$$

where ℓ is the length of diagonal structures larger than $\ell_{\min}$ (in this study we use $\ell_{\min} = 2$). From a heuristic point of view, high DET values are related to higher predictability. Although there are further appropriate RP measures, in this study, to detect regime transitions, we focus on DET only.

III. MODEL EXAMPLES

In real-world applications, such as palaeoclimate studies, data sets are normally irregularly sampled due to the nature of the nonlinear decomposition rates. Before we apply the proposed metric distance to such an example from palaeoclimate research, we demonstrate the potential of the approach using prototypical models.

A. Logistic map

As a first application we choose the one-dimensional discrete logistic map, which is a simple mathematical model to mimic population dynamics given by

$$x_{i+1} = r x_i (1 - x_i) \tag{6}$$

for the control parameter $r \in [0, 4]$ [22]. We examine the system in the range $r \in [3.5, 4]$ where the logistic map possesses rich dynamics: accumulation points, periodic and chaotic behaviors, merging of chaotic bands, period doublings, inner and outer crises, etc. Abrupt transitions between these regimes occur by varying r.

First we generate a regularly sampled time series with Eq. (6). The system is iterated for 3000 points for each value of $r \in [3.5, 4]$ ($\Delta r = 0.01$). Two thousand points are truncated to discard transients, resulting in regularly sampled time series with 1000 points for each r value. To mimic irregular sampling, we randomly delete a certain fraction γ of points. We prepare four different irregularly sampled data sets by randomly deleting 50 ($\gamma = 5\%$), 100 ($\gamma = 10\%$), 150 ($\gamma = 15\%$), and 200 ($\gamma = 20\%$) points from the original time series.

For the four irregularly sampled data sets, we create RPs (for the range of the parameter r) according to Eq. (4) and using an arbitrarily chosen sequence size ensuring that approximately six data points fall into the sequence. The parameters $\Lambda_{0,k}$ are calculated using Eqs. (2a) and (2b). We find that choosing $\Lambda_S = 1$ gives a smooth bell-shaped frequency distribution for the distances D. As mentioned, the RP threshold is selected as $\varepsilon = 2\sigma_D$. Finally, the different dynamical regimes according to r are detected by calculating the DET values for each RP.

In order to quantify the performance of our approach for different levels of perturbations by deleting points, we also compute the Lyapunov exponent λ, which is a measure of divergence of infinitesimally close trajectories [23] and DET [Eq. (5)] of the original time series (Fig. 2). The Lyapunov exponent is a well-known and powerful measure to understand the behavior of complex systems. For chaotic dynamics the Lyapunov exponent is positive, whereas for periodic dynamics it is negative. When the Lyapunov exponent hits zero from negative and turn back to the negative region, the system has a period-doubling behavior, i.e., a transition from a periodic case to another periodic case. If the Lyapunov exponent crosses the zero line, then the system has a critical transition from a periodic to a chaotic regime or vice versa (dotted lines in Fig. 2). By using our approach, DET is able to detect those transitions even in the case of missing points (mimicking irregular sampling) [Fig. 2(b)]. The amplitudes of change in DET decrease when more data points are removed, because deleting points basically breaks diagonal lines by this operation. However, we can still clearly detect all important regime changes.

Another important issue in time-series analysis is noise. Therefore, we extend this example of the logistic map by adding noise and test the performance of our method against different levels of noise. Gaussian white noise is added with noise levels of $0.05\sigma(x)$ and $0.1\sigma(x)$, where $\sigma(x)$ is the standard deviation of the time series, on the data of $\gamma = 5\%$ and 10% randomly deleted cases for the irregular sampling. The recurrence plot threshold is $\varepsilon = 2\sigma_D$. The variation of DET indicates most of the regime transitions, despite the noise, although some of the smaller changes in DET are not as clear as without noise [Fig. 2(c)]. These results show that we can still detect all transitions in the presence of additional difficulties.

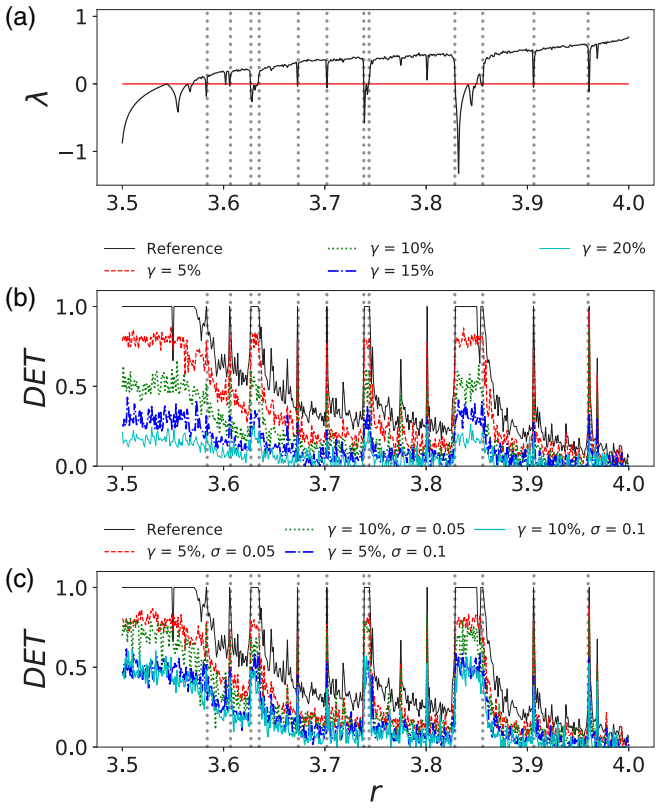

FIG. 2. Recurrence analysis of the logistic map: (a) Lyapunov exponent λ, (b) DET for reference series and time series with randomly deleted points, and (c) DET for reference series and time series with randomly deleted points with additive noise. Dotted vertical lines show regime transitions.

The metric distance D is introduced and used to construct TACTS [5]. In TACTS, we only measure the distance of consecutive data segments S_i, where $i = 1, \ldots, \omega$, and it creates a vector (single-variate time series), i.e., TACTS is a kind of difference detrending filter. In the present study we create a matrix by computing the distances, using the same metric distance measure, between all pairs of ω sequences. Thresholding this matrix constructs a recurrence plot. This approach is different from the TACTS approach, because here we do not consider a difference filtered time series but pairwise distances (similarities) based on the local transformation costs. For comparison, we perform a similarity analysis between RQA and the Lyapunov exponent with the TACTS approach and the current approach. When we conduct the similarity analysis, the Pearson correlation coefficient is used,

$$\rho_{(x,y)} = \frac{\mathrm{cov}(x, y)}{\sigma_x \sigma_y}, \tag{7}$$

TABLE I. Similarity-based comparison of TACTS [5] and the approach by using absolute value of the Pearson correlation of the Lyapunov exponent λ and the DET values $|\rho_{(\lambda,\mathrm{DET})}|$.

$\lvert\rho_{(\lambda,\mathrm{DET})}\rvert$	reference	$\gamma = 5\%$	$\gamma = 10\%$	$\gamma = 15\%$	$\gamma = 20\%$	$\gamma = 5\%, \sigma = 0.05$	$\gamma = 5\%, \sigma = 0.1$	$\gamma = 10\%, \sigma = 0.1$	$\gamma = 10\%, \sigma = 0.1$
TACTS	0.89	0.87	0.86	0.87	0.87	0.86	0.85	0.83	0.85
RP with D	0.90	0.90	0.89	0.87	0.80	0.90	0.89	0.89	0.89

where $\mathrm{cov}(x, y)$ is the covariance of two time series x and y, and σ_x and σ_y are their standard deviations, respectively. The results are illustrated in Table I.

Except for the case of $\gamma = 20\%$, the matrix approach always has higher correlation than the TACTS method. As is well known, high-pass filters are not very efficient preprocessing techniques for noisy time series. However, our approach is rather robust against noise while TACTS is affected by it (see Table I).

B. Rössler attractor

After the test of our approach on a discrete system, we now move to a paradigmatic continuous system, offering more testing opportunities. We consider the Rössler attractor [24]

$$\left(\frac{dx}{dt}, \frac{dy}{dt}, \frac{dz}{dt}\right) = (-y - z,\ x + ay,\ b + z(x - c)). \quad (8)$$

The system has three control parameters a, b, and c. Similar to the logistic map, we vary one control parameter to have regime transitions, i.e., $b \in [0, 1.4]$ with a resolution of $\Delta b = 0.01$ and with constants $a = 0.2$ and $c = 5.7$. First we generate time series for each control parameter value b by integrating Eq. (8) using a fourth-order Runge-Kutta scheme and with $\Delta t = 0.01$ sampling time. To mimic irregularly sampled time series, we select 5000 maximum points from the y component of the continuous time series [Fig. 3(a)]. This is a natural way to achieve irregularly sampled time series from continuous time series [5].

In palaeoclimate data, the sampling times often follow a Gamma distribution [4]

$$p(k, s) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/s}. \quad (9)$$

Therefore, we create different time series by interpolating to sampling times that follow a Γ distribution with different parameters $s = 0.3$, 0.5, 1.0, and 2.0 (Fig. 3). Clearly, the uncertainty of the time series increases as the skewness increases.

Next we follow the same procedure as in the logistic map application.

(i) Divide time series into small sequences of size ensuring that approximately six points fall inside, $\frac{6}{N-1}\sum_{i=1}^{N-1} t_{i+1} - t_i$, for all values of b.

(ii) Calculate D for all sequence pairs by using Eqs. (1), (2a), and (2b), using $\Lambda_S = 1$ as in the logistic map.

(iii) Calculate RPs with $\varepsilon = 2\sigma_D$.

(iv) Calculate DET from the RPs.

(v) Calculate the Lyapunov exponent λ from the continuous time series.

For the considered range in the parameter b, several regime transitions between chaos and periodic (or periodic to chaos) occur (Fig. 4, dotted vertical lines). These transitions are clearly detectable with the proposed approach for the undisturbed time series, but also for the irregularly sampled cases. Although increasing skewness s lowers the changes in DET, they are still easily observable. Therefore, our approach is still valid for highly skewed or perturbed data sets.

IV. APPLICATION TO PALAEOCLIMATE RECORDS

To illustrate the applicability of our method we select a recently published speleothem proxy time series from the Dayu Cave (DY1) [8], which is under the influence of the East Asian Summer Monsoon (EASM). In comparison to

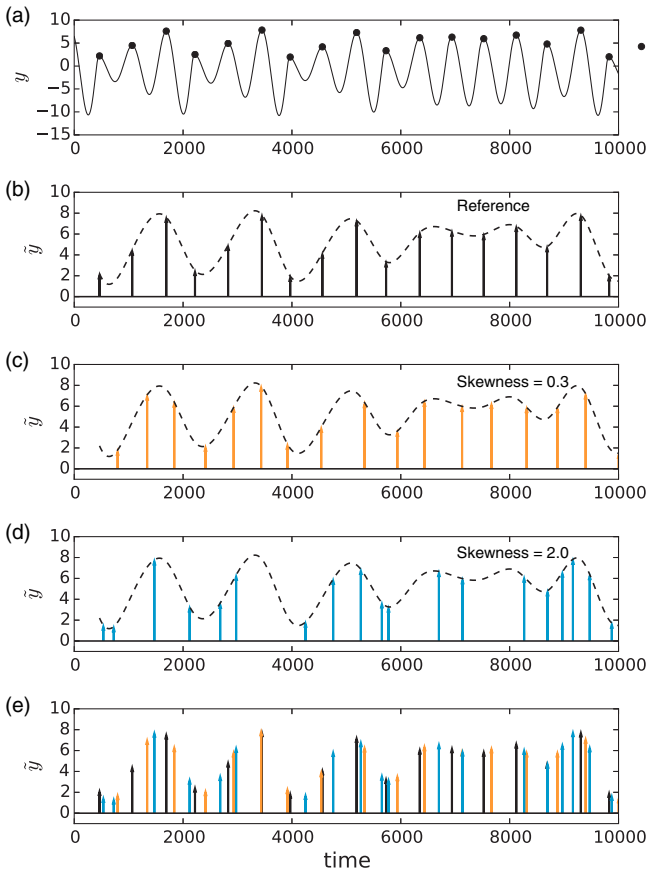

FIG. 3. Extracting data from the Rössler attractor. (a) Maxima of the y component of the time series highlighted as bullets, (b) result of the maximum map $\tilde{y}$ acting on y, (c) time points after resampling or interpolation using the Γ function with skewness $s = 0.3$, (d) same as (c) but for $s = 2.0$, and (e) comparison of the irregular sampling points as derived from the different skewness levels (b)–(d). In (b)–(d) the dashed line represents the interpolated time series corresponding to the maxima in (a).

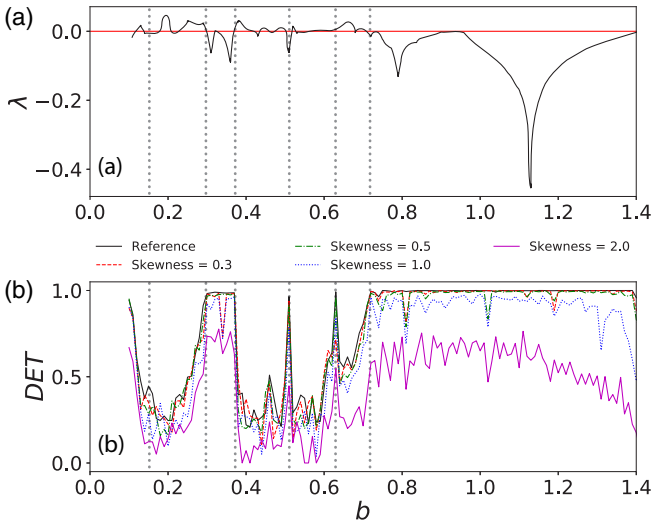

FIG. 4. Recurrence analysis of the Rössler oscillator: (a) Lyapunov exponent λ and (b) DET for reference series and the Γ-distributed time series with skewness $s = 0.3$, 0.5, 1.0, and 2.0. Dotted vertical lines show regime transitions.

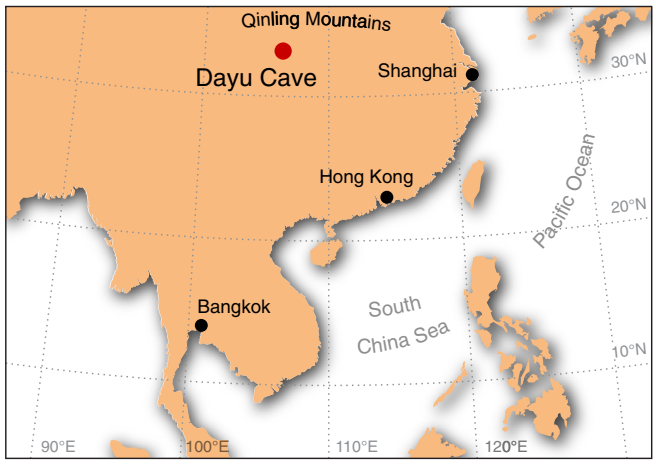

FIG. 5. Location of the Dayu Cave at the southern rim of the Qinling Mountains.

other palaeoclimate archives like lake sediments or ice cores, speleothems frequently offer the advantage of highly robust U-series chronologies in conjunction with long growth periods without diagenetic alterations [25]. Among the available speleothem records, the δ^{18}O time series from Dayu Cave is exceptional as its sensitivity to changes in local monsoonal rainfall is corroborated by historic graffiti inscriptions found on the cave's walls [8]. The excellent match between drought events recorded in these graffiti and significant excursions in the isotope time series provides a rigorous test for the indirect proxy information. While a distal moisture source, increased rainfall, prolonged infiltration, and reduced reevaporation, all associated with strong summer monsoon periods, lead to negative δ^{18}O values, weak EASM episodes are recorded in the stalagmite as positive δ^{18}O excursions [8,26,27]. Besides such more linear and obvious changes in the δ^{18}O record, more subtle changes related to changing dynamics of the monsoon system (e.g., change in regularity) are not obvious and require more advanced methods such as considering recurrences [28]. A correct detection of abrupt or gradual changes and extreme events in long palaeoclimate reconstructions is of great interest as they serve as test cases for the influence of climate dynamics on society [29–31]. At the same time, chronological uncertainties inherent to such time series often complicate the extraction of information on short-lived (months to years) extreme events or change points.

Dayu Cave is located on the southern slopes of the Qinling Mountains, central China (Fig. 5). The region is highly susceptible to changes in summer monsoon strength as it receives most moisture during the EASM season between June and October. The proxy time series from stalagmite DY1 covers approximately the past 700 years with 550 samples. Growth rate changes result in irregular temporal resolution (median sampling time 1.31 yr, 5%-quantile 0.65 yr, and 95%-quantile 2.01 yr). The chronology of the δ^{18}O time series is based on 15 U-series dates with maximum 2σ error of 8 yr (see [8] for details).

In order to detect regime transitions in the palaeoclimate time series, we apply a sliding window of 120 yr with a step of 10 yr. The sliding windows are divided into small sequences with a size of 4 yr to compute RPs with the metric distance. For metric distance calculation, the parameters Λ_0 and Λ_k are computed by using Eq. (2a) and (2b), respectively. The adding and deleting cost Λ_S is adaptively chosen as 1.02. For consistency of the RPs, the threshold ε is fixed to $\varepsilon = 2\sigma_D$.

The δ^{18}O time series and associated DET values are given in Fig. 6. The confidence interval is computed by a bootstrapping approach to detect significant changes in the evolution of

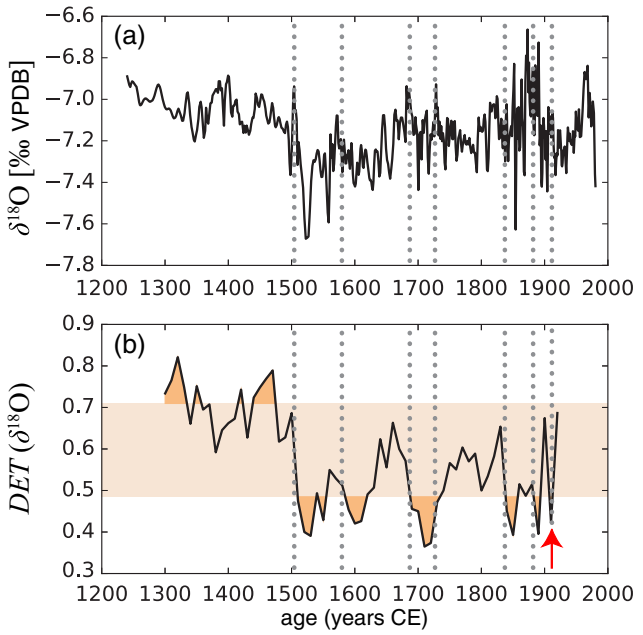

FIG. 6. (a) The δ^{18}O record from the Dayu cave and (b) variation of the DET measure for the δ^{18}O record. Dotted lines represent documented drought events and shaded area in (b) denotes the confidence interval for DET.

DET [32]. If the DET curve is within the confidence interval, the system is in "normal conditions" (90% confidence). When the curve exceeds the confidence interval, we consider it a significant regime transition.

Eleven significant events are detected by our approach and lower DET values meet with the drought events mentioned by historical records in 1528, 1596, 1707, 1756, 1839, 1891, and 1894 [8] [Fig. 6(b)]. Although ancient people left no inscription in Dayu Cave at the time, we detected another event for 1908 AD, which corresponds to a significant historical drought in the region [33] [Fig. 6(b), arrow]. Interestingly, DET seems very sensitive to local rainfall conditions. In nearly all cases DET values begin to decrease with, or slightly before, the onset of severe droughts and return to higher values when droughts abate. More work with additional high-resolution time series is needed to confirm this finding.

The DET measure abruptly shifts from high values to lower ones about 1500. This time marks the onset of a period with extremely cold winters, which continued to 1900, and corresponds the Little Ice Age (LIA). The LIA was characterized by an intensified winterly Siberian High, which brings cold dry air from northerly directions to China [34,35]. It is worth mentioning that the DET values are always low throughout this period. Earlier studies made two interesting observations: (i) Higher DET values indicate a stronger monsoon and otherwise weak monsoon activity and (ii) the monsoon is driven by solar activity [36,37]. These findings are confirmed by our results; we detect that low DET values correspond closely with drought events and the Spörer-Maunder solar minima [38]. Although the variation of DET gives results consistent with previous findings, we do not expect that this is a general rule. For distinct regions, monsoon dynamics can have different dynamical behavior. Therefore, drought times are not necessarily linked with less predictable dynamics (or strong monsoon periods might not always be linked with more predictable dynamics) globally. Here we only note that in the considered proxy the dynamics of the monsoon activity follows the mentioned characteristic pattern.

Prior to 1500 we observe high DET values [Fig. 6(b)], which might be indicative of either strong EASM periods or alternatively simply enhanced seasonality, with both stronger EASM rainfall and simultaneously intensified drought during the dry seasons. Two of the detected events (1320 and 1420) correspond to stronger monsoon activity, whereas the other events seem to be related to droughts in 1350 and 1470 [39]. It remains unclear therefore if high and low DET values correspond to strong shifts in climate dynamics alone or could possibly signify enhanced seasonality and wet and dry periods, respectively. This should be tested using longer time series from different regions characterized by strong seasonality.

Our approach successfully detects nine known events out of eleven. It is a very interesting question why our approach only identifies two wet intervals as corresponding to high DET values. Additional palaeoclimate research is required to give insights into local effects that could affect the sensitivity of the proxy time series to regional climate dynamics; however, such work is beyond the scope of this paper.

V. CONCLUSION

Analyzing time series with irregular sampling is a challenging task and often requires data preprocessing, e.g., interpolation. Here we proposes an approach that allows analyzing irregularly sampled time series in particular for methods which are based on a distance metric, such as recurrence plots. The application of the transformation cost time series approach as a metric for the construction of RPs allowed us to identify dynamical differences or to detect regime transitions. We have demonstrated the potentials on discrete and continuous nonlinear prototypical systems. Testing this approach on a specific palaeoclimate reconstruction highlights the practical applicability. Using our approach, we were able to identify characteristic regime transitions in the regional climate during the past 700 years that are known to have caused droughts or are related to the Little Ice Age. This method is useful on "imperfect" (i.e., irregularly sampled) time series. If the data are regularly sampled, the method is not required. In this work, we analyzed a single disparate palaeoclimate reconstructions in order to gain insights into spatiotemporal variability of key climate parameters. This technique not only is helpful for recurrence analysis of irregularly sampled time series, but can also be used for any other distance-based methods. The distance approach not only is helpful for recurrence analysis of irregularly sampled time series, but can also be used for any other method which is based on distances.

ACKNOWLEDGMENTS

We acknowledge support from the DFG through Project No. MA4759/9-1: Recurrence plot analysis of regime changes in dynamical Systems; from TUBITAK, 2214/A program; from Ege University, Research Project No. 2015FEN028; and from the European Union's Horizon 2020 Research and Innovation program under the Marie Skłodowska-Curie Grant Agreement No. 691037.

[1] J.-P. Eckmann, S. Oliffson Kamphorst, and D. Ruelle, Europhys. Lett. **4**, 973 (1987).
[2] P. Grassberger, Phys. Lett. A **97**, 227 (1983).
[3] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, Phys. Rev. Lett. **45**, 712 (1980).
[4] K. Rehfeld, N. Marwan, J. Heitzig, and J. Kurths, Nonlinear Process. Geophys. **18**, 389 (2011).
[5] I. Ozken, D. Eroglu, T. Stemler, N. Marwan, G. B. Bagci, and J. Kurths, Phys. Rev. E **91**, 062911 (2015).
[6] N. Marwan, M. C. Romano, M. Thiel, and J. Kurths, Phys. Rep. **438**, 237 (2007).
[7] S. Suzuki, Y. Hirata, and K. Aihara, Int. J. Bifurcat. Chaos **20**, 3699 (2010).
[8] L. Tan, Y. Cai, Z. An, H. Cheng, C.-C. Shen, S. F. M. Breitenbach, Y. Gao, R. L. Edwards, H. Zhang, and Y. Du, Sci. Rep. **5**, 12284 (2015).
[9] J. J. C. Michael T. Rosenstein, and C. J. D. Luca, Physica D **65**, 117 (1993).

[10] T. Pereira, D. Eroglu, G. B. Bagci, U. Tirnakli, and H. J. Jensen, Phys. Rev. Lett. **110**, 234103 (2013).
[11] D. Eroglu, J. S. W. Lamb, and T. Pereira, Contemp. Phys. **58**, 207 (2017).
[12] A. Groth, Phys. Rev. E **72**, 046220 (2005).
[13] K. Rehfeld, N. Marwan, S. F. M. Breitenbach, and J. Kurths, Climate Dyn. **41**, 3 (2012).
[14] J. D. Victor and K. P. Purpura, Network: Comput. Neural Syst. **8**, 127 (1997).
[15] D. Eroglu, T. K. DM. Peron, N. Marwan, F. A. Rodrigues, L. d. F. Costa, M. Sebek, I. Z. Kiss, and J. Kurths, Phys. Rev. E **90**, 042919 (2014).
[16] D. Eroglu, N. Marwan, S. Prasad, and J. Kurths, Nonlinear Process. Geophys. **21**, 1085 (2014).
[17] L. Kabiraj, A. Saurabh, H. Nawroth, and C. O. Paschereit, AIAA J. **53**, 1199 (2015).
[18] H. Poincaré, Acta Mathematica **13**, 1 (1890).
[19] N. Marwan, Int. J. Bifurcat. Chaos **21**, 1003 (2011).
[20] S. Schinkel, O. Dimigen, and N. Marwan, Eur. Phys. J. Spec. Top. **164**, 45 (2008).
[21] L. L. Trulla, A. Giuliani, J. P. Zbilut, and C. L. Webber, Jr., Phys. Lett. A **223**, 255 (1996).
[22] R. M. May, Nature (London) **261**, 459 (1976).
[23] A. Wolf, J. B. Swift, H. L. Swinney, and J. A. Vastano, Physica D **16**, 285 (1985).
[24] O. Rössler, Phys. Lett. **57**, 397 (1976).
[25] I. Fairchild and A. Baker, *Speleothem Science: From Process to Past Environment* (Wiley-Blackwell, Chichester, 2012).
[26] S. F. Breitenbach, J. F. Adkins, H. Meyer, N. Marwan, K. K. Kumar, and G. H. Haug, Sci. Lett. **292**, 212 (2010).
[27] A. Baker, J. C. Hellstrom, B. F. J. Kelly, G. Mariethoz, and V. Trouet, Sci. Rep. **5**, 10307 (2015).
[28] N. Marwan and J. Kurths, Chaos **25**, 097609 (2015).
[29] S. Brönnimann, J. L. Annis, C. Vogler, and P. D. Jones, Geophys. Res. Lett. **34**, L22805 (2007).
[30] D. J. Kennett, S. F. M. Breitenbach, V. V. Aquino, Y. Asmerom, J. Awe, J. U. L. Baldini, P. Bartlein, B. J. Culleton, C. Ebert, C. Jazwa *et al.*, Science **338**, 788 (2012).
[31] W. Steffen, J. Rockström, K. Richardson, T. M. Lenton, C. Folke, D. Liverman, C. P. Summerhayes, A. D. Barnosky, S. E. Cornell, M. Crucifix *et al.*, Proc. Natl. Acad. Sci. USA **115**, 8252 (2018).
[32] N. Marwan, S. Schinkel, and J. Kurths, Europhys. Lett. **101**, 20007 (2013).
[33] L. Tan, Y. Cai, L. Yi, Z. An, and L. Ai, Clim. Past **4**, 19 (2008).
[34] J. J. A. Matthews and K. K. R. Briffa, Geogr. Ann. A **87**, 17 (2005).
[35] Q. Ge, J. Zheng, Z. Hao, Y. Liu, and M. Li, J. Geogr. Sci. **26**, 827 (2016).
[36] D. Eroglu, F. H. McRobie, I. Ozken, T. Stemler, K.-H. Wyrwoll, S. F. M. Breitenbach, N. Marwan, and J. Kurths, Nat. Commun. **7**, 12929 (2016).
[37] F. A. Lechleitner, S. F. M. Breitenbach, K. Rehfeld, H. E. Ridley, Y. Asmerom, K. M. Prufer, N. Marwan, B. Goswami, D. J. Kennett, V. V. Aquino *et al.*, Sci. Rep. **7**, 45809 (2017).
[38] J. A. Eddy, Science **192**, 1189 (1976).
[39] J. Zheng, W. C. Wang, Q. Ge, Z. Man, and P. Zhang, Terr. Atmos. Ocean. Sci. **17**, 579 (2006).